DOCUMENT RESUME

ED 299 326 TM 012 371

AUTHOR Arrasmith, Dean G.; Hambleton, Ronald K.
TITLE Steps for Setting Standards with the Angoff Method.
PUB DATE 6 May 88
NOTE 26p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Certification; Credentials; *Licensing Examinations (Professions); Minimum Competencies; *Minimum Competency Testing; Multiple Choice Tests; *Standards
IDENTIFIERS *Angoff Methods; *Standard Setting

ABSTRACT

Specific steps for applying the Angoff method are described. In the Angoff method, judges are asked to estimate the probabilities of minimally competent candidates' answering multiple choice test items correctly. Initial information must be obtained for designing the standard-setting process, beginning with the purpose of the examination and any restrictions imposed. In preparation for the standard setting meeting, four areas of concern must be addressed: (1) background for certifying; (2) meeting site selection; (3) judge selection; and (4) preparation of materials for judges. At the meeting, three main areas must be considered: (1) examination background; (2) definition of an entry-level professional; and (3) setting the standard. Setting the standard requires a practice exercise and review of the item rating process. A post-standard setting meeting is required to document the process and complete documentation. The Task Review Form, the Angoff Item Review Forms (Methods A and B), and an implementation checklist for the Angoff Standard-Setting Method are appended. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299326

STEPS FOR SETTING STANDARDS WITH THE ANGOFF METHOD[1,2]

Dean G. Arrasmith and Ronald K. Hambleton
University of Massachusetts, at Amherst

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ronald K. Hambleton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012 371

05/06/87

Steps for Setting Standards with the Angoff Method[1,2]

Dean G. Arrasmith and Ronald K. Hambleton
University of Massachusetts, at Amherst

A critical problem in the establishment of a valid credentialling examination program concerns the setting of the standard or passing score. The purpose of this report is to describe specific steps for applying the well-known and popular Angoff method (Angoff, 1971; Bernknopf, Curry, & Bashaw, 1971; Livingston & Zieky, 1982). In the Angoff method, judges are asked to estimate the probabilities of minimally competent candidates answering multiple-choice test items correctly. However, if the standard-setting process is to be defensible to the candidates, to the public, and in the courts, many steps, in addition to the one where judges estimate performance levels of minimally competent candidates, must be carried out correctly and documented. In this report, attention will be focused on the complete set of steps and on the implementation of the steps.

In reviewing the Angoff method, van der Linden (1982) identified three important sources of inconsistency in judges' ratings (1) different conceptions of minimal proficiency or mastery, (2) different interpretations of the content measured by the exam, and (3) carelessness and/or shifts in the use of rating scales during the item rating process. In this report, suggestions are offered to overcome

---

[1] Support for the preparation of this report was provided by Professional Examination Service.

[2] Laboratory of Psychometric and Evaluative Research Report No. 168. Amherst, MA., School of Education, University of Massachusetts.

PES.7

the problems van der Linden described. In addition, Hambleton and Powell (1983) and Popham (1986) described a large number of factors which should be considered in order to set defensible standards. For example, Hambleton and Powell considered context variables such as importance of the exam, laws, availability of resources and exam formats, content, and length. These, and other factors are addressed in the report. Our hope is that the report will help groups (1) choose a standard setting process that addresses all of the steps outlined in the report and (2) implement each step fully and with complete documentation.

The steps for setting defensible standards are organized into four broad sequential categories of issues/guidelines for effectively implementing the Angoff method (or variations):

I. Initial Information
II. Standard-Setting Meeting Preparation
III. Standard-Setting Meeting
IV. Post Standard-Setting Meeting

Each of the categories above will be discussed subsequently, and in a final section of the report, a checklist for use in both the design and/or evaluation of the standard-setting process is presented and discussed. Category I is intended to provide the essential background information about the exam prior to choosing and implementing a standard-setting process. Categories II, III, and IV pertain to standard setting activities which arise (1) before the standard-setting meeting, (2) during the meeting, and (3) following the meeting. Note that the defensibility of the standard-setting

process will involve a defense of the activities carried out in each category not just a defense of the actual judges' item rating task.

I. Initial Information

In getting started with the standard-setting process, six main questions should be addressed by the individual or organization responsible for designing the standard-setting process. These six questions should be answered in written form to leave a history of the background information that was used in choosing a standard-setting process. Rationales for decisions should be prepared along with a list of individuals who were involved in the decisions. Such a history can help describe and defend the standard to the candidates and/or the public. Here are the questions:

1. What is the purpose for the examination?
   (Briefly describe the purpose of the exam and the content measured by the exam.)

2. How important are the results from the exam to candidates, institutions, or to the public interest? (Be sure to consider the potential gain and potential harm the exam has for each group. Also consider if the exam scores are used with other sources of information about the candidates when final decisions are made. Consider too whether candidates can retake the exam, and if so, consider the frequency.)

3. Can the exam be used for its intended purpose(s) without setting a standard?
   (Determine whether pass/fail decisions are required to fulfill the purpose for the exam. Would exam performance, e.g., number or percentage correct, offer sufficient information about examinee performance?)

4. Are there significant limitations on time, money and resources, candidates for pilot testing, and availability of qualified judges? (These limitations must be reconciled with the purpose and importance of the exam. If the exam is very important then adequate time, money, resources and judges should be found or identified.)

5. Are there significant regulations or restrictions that influence the exam, the standard setting process, the standard, or the decisions about candidates? For example, often standards must be set separately for each state using an examination.
(Consider the requirements about the exam and standard setting process that may be imposed by mandates or political and legal necessity. Such requirements may include the specific exam content and/or length; the judges who must participate in the standard setting process; the standard, specified or constrained in some manner; and the number and demographics of examinees passing and failing the exam. In addition, consideration should be given to the time and preparation candidates have been given to be ready for the examination.)

6. How was the exam developed?

   A. How was the exam content identified?

      a. Was a role delineation study conducted?
      b. Is there a set of domain specifications?
      c. Is there a list of behavioral objectives?
      d. Is there an exam blueprint? How was it prepared and by whom?

   B. Who reviewed the exam content? (Identify the individuals and their roles. Such information may be needed later to defend the choice of exam content.)

   C. How was the exam constructed?
   (Describe how exam items were obtained or written. Who wrote the exam items? What information did the writers have? Who reviewed and how did they review the exam items? If multiple forms are available what evidence is there that the forms are parallel?)

   D. What item types were included in the exam?
   (e.g. multiple-choice, multiple true-false, matching)
   (Item format information is important because the Angoff method will only be appropriate for some item types, and must be modified for others.)

If the questions above have been carefully and fully answered, there is a history of test purpose and development that will be useful for documentation purposes and in selecting appropriate steps for implementing the Angoff standard-setting process.

II. Standard-Setting Meeting Preparation

Prior to the actual standard-setting meeting with judges, four areas of concern must be addressed: A. Background, B. Meeting Site Selection, C. Judge Selection, and D. Preparation of Judges' Packages.

A. Background

In setting standards for credentialling exams, the results from a role delineation study are usually available to guide the selection of exam content. However, in some specialties, particularly the newer and/or less well-developed specialties, domain specifications or behavioral objectives may be available. Another variation is to describe exam content by a two-dimensional grid where the dimensions may correspond to major content topics and levels of thinking (i.e. knowledge, comprehension, application, etc.) . Steps corresponding to the format of the content (role delineation, domain specifications, objectives, or exam blueprint) described below should be followed. Choose one of the four options below to carry out step one:

Role Delineation Results Available

1. (a) Prepare the role delineation statement for the judges. (Includes responsibilities, sub-responsibilities, activities, skills and knowledge statements.)

   (b) Determine the appropriate level in the role delineation statement on which to focus judges' attention when setting the standard. (Normally a role statement includes 5 to 10 responsibilities, 20 to 40 sub-responsibilities, and over 100 activities, knowledge, and skill statements. For the purposes of setting a standard, usually it is sufficient to focus judges on 20 to 40 tasks in the role. In the typical case then, judges' attention should be focused at the sub-responsibility level.)

(c) Code the exam items by the sub-responsibility they measure.

Domain Specifications Available

1. Code exam items by the domain specification numbers they measure. (Regroup the domain specifications into 20 to 40 statements if the initial list exceeds 20 to 40.)

Objectives Available

1. Code exam items by the objective numbers they measure. (Regroup the objectives into 20 to 40 statements if the initial list exceeds 40.)

Test Blueprint Available

1. Prepare a copy of the exam blueprint for distribution to judges. Code the exam items to the cells in the exam blueprint.

We note that it is often useful to provide judges with the content specifications prior to the standard-setting meeting so that they will have time to read them in detail. Often the exam content information is quite long.

In addition to the matching of items to the content specifications for the exam, two additional activities should be carried out in this phase of the process:

2. Prepare a scoring key.

3. Compile exam item statistics (if available). (Item difficulty, discrimination, and distractor effectiveness information can be of considerable value to judges in the standard-setting process.)

B. Meeting Site Selection

Inclusion of this category may seem unusual; however, our experience has been that the quality of work is considerably better

when careful attention is given to the meeting site selection and associated work space.

1. Choose a suitable location for the meeting. (The meeting should be held in a place that is convenient for the judges to attend and is free from distraction for the judges. If cost and inconvenience are a concern, multiple-site meetings may be necessary.)

2. Insure sufficient work space is available for the meeting. (Judges will need a. work tables to complete the several rating forms, b. space for work-group [3-4 people] meetings, and c. space for the entire group to meet.)

3. Arrange for coffee service and lunch (optional).

4. Consider the incentive to judges of a nice retreat site.

C. Judge Selection

1. Identify the desired and required demographic and technical characteristics of the judges. (Consider for inclusion, for example, appropriate ethnic, gender and age groups, content experts, workers, national leaders, and the general public served.)

2. Determine the number of judges to be selected to meet each desired characteristic. (Inherent in the choice of number of judges will be the importance to the project and relative size of the populations represented by the judges. When possible, judges should represent more than one constituency, e.g. a content expert who is Hispanic. Ten to 20 judges for credentialling exams is usually a desirable number; however, in some instances somewhat smaller numbers may be all that is feasible. Somewhat more judges are normally desirable with educational tests.)

3. Determine if compensation to the judges is necessary or even possible. (Some or all judges may need to be compensated for their time. More usually however, judges and their employers are willing to contribute the time involved in standard-setting to enhance the profession.)

4. Select judges.
   (Every effort should be made to locate a representative sample of judges meeting the required composition defined in step 1 above.)

5. Organize judges into work groups before their arrival to save time and reduce mistakes. (Be prepared however, to find replacements for judges who can't attend the meeting, or to make group reassignments when the numbers do not meet the goal.

   Each work group should be chosen to be representative of the entire group. In this way, independent replications of the standard-setting process can be set up and the resulting standards compared. Normally groups of 3 to 5 judges are workable. Three to four groups are also useful.)

D. Preparation of Judges' Packages

Packages with the material below should be prepared for judges:

1. Role statement (or domain specifications, objectives, or exam blueprint)
2. Exam items (coded to the content they measure)
3. Answer key (optional)
4. Item statistics (if available)
5. Angoff Item Rating Form
6. Pencil

We note that it is often convenient for the purpose of providing Angoff item ratings to have the exam items organized into groups which correspond to sub-responsibilities, objectives, or cells of the exam blueprint. However, reorganizing the exam items can be time-consuming to set up, and mistakes can be made. Alternately and conveniently, exam items in the exam booklet can be coded to the content they measure.

III. Standard-Setting Meeting

In this category, there are three main areas of concern that must be addressed: A. Background, B. Definition of an Entry-level Professional, and C. Setting the Standard.

(The information from Category I will be especially helpful in providing details for the following steps.)

A. Background

1. Introduce the order of work for the judges (see A2 to A4, and sub-categories B and C on the next few pages)

2. Describe the history of the exam development project (purpose and development)

3. Present the rationale for setting the standard (or standards)

4. Stress the importance of the exam and the standard to the profession

B. Definition of Entry-level Professional

There are two variations on the implementation, denoted Methods A and B. Method A is recommended but Method B requires less time.

Method A

(The general strategy adopted here is to first focus judges on the content of the exam reflected in the role delineation, domain specifications, objectives, or exam blueprint. Then to reach a joint understanding of what entry level competence is.)

1. Have each judge review the role delineation statement, domain specifications, or objectives.

2. Instruct the judges in the use of the Task Review Form. (See Appendix A)

3. Have each judge complete the Task Review Form. (See Appendix A)

4. As a group, have the judges decide on a compromise (or average) rating (percent) for each task or objective. Prior to averaging the ratings, provide the judges with the opportunity to discuss their ratings. Focus attention on the extreme ratings in the discussion (highest and lowest).

5. Compile a final listing of ratings (percent) for each task (objective). (This listing is a profile of a minimum-competent candidate over the content measured by the exam.)

6. Add the ratings (percent) to the "Angoff Item Review Forms." (See Appendix B)

Method B

(The general strategy adopted here is to first review the general content of the examination with the judges. Then, through group discussion with the judges, operationally define a description of the entry level professional.)

1. Have each judge review the exam blueprint or other description of the content of the exam.

2. Initiate a discussion about the necessary skills for an entry-level professional. (A good way to encourage and focus this discussion is to ask for necessary skills within each cell (category) of the exam blueprint. The outcome of this task is a general listing of desirable skills.)

3. Review the list of skills with the judges. (An overhead projector would be useful for presenting this list to the group. The objective of this review is to (a) eliminate redundancies, (b) reject any inappropriate skills and (c) check for completeness of the list.)

4. If possible, copy and distribute the list of necessary skills for an entry-level professional. (Again, it is convenient to organize this list by the cells (categories) in the exam blueprint.

C. Setting the Standard

(Focusing on the profile developed in Method A or B above, the judges will estimate the difficulty of each exam item for the minimally competent candidate.)

1. Work with the judges on a practice exercise. (Ten to 20 exam items should be sufficient.)

2. Answer any questions about the item rating process. (Be sure to emphasize the two stages of review and the need to work as a group in the final review stage.)

3. Distribute the Angoff Item Review Forms (see Appendices B and C), exam items and answer keys.

4. Separate the judges into work groups.

5. Have the judges complete the item reviews.

6. Have judges within each group discuss their ratings and revise their ratings when necessary. Attention should be focused on items where judges disagree in their percent assignments in excess of 20%.

7. Each group member should aggregate his/her revised ratings to produce an individual standard and then the individual standards should be averaged to produce a single group standard.

8. Bring the groups together and then average the standards from each work group to produce a single standard.

9. Review and discuss the standard with the judges. Work groups will want to explain why their standards are higher or lower than the average standard.

If item statistics are available, they may be useful to judges in providing a framework for providing their ratings. Knowing that (say) 30% of the candidates answered a question correctly can be useful in setting expected performance levels for minimally proficient candidates.

When a standard is applied to actual exam results, there are two kinds of errors: false-positive and false-negative. If false-negative errors are viewed as far more serious than false-positive errors, the standard can be lowered by one, two, or three standard errors of measurement to decrease the likelihood of making false-negative errors. On the other hand, the lower the standard is set, the more the number of false-positive errors is increased. Alternately, the standard can be increased by one, two, or three standard errors of measurement when the goal is to minimize the number of false-positive errors. Correspondingly, the number of false-negative errors is increased.

IV. Post Standard-Setting Meeting

A. Documentation of the Process

1. Prepare an outline of the major steps taken; provide rationales for the actions whenever possible.

2. Note any problems, inconsistencies or threats to the validity of the standards.

B. Distribution of the Final Documentation

1. Send copies to judges for review and for their records (optional).

2. Place a copy on file for future reference and use.

Summary/Implementation Checklist

In order to summarize the tasks required to implement the Angoff standard-setting method, a checklist of tasks has been compiled and is contained in Appendix D. This checklist should be useful in planning and controlling the implementation of the Angoff method. It is suggested that the date of completion of each task be recorded. In this way, the checklist should be useful in guiding the implementation, and helpful in suggesting appropriate tasks.

References

Angoff, W. (1971). Scales, norms, and equivalent scores. In R. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education.

Bernknopf, S., Curry, A., & Bashaw, W.L. (1979, April). A defensible model for determining a minimal cut-off score for criterion-referenced tests. Paper presented at the annual meeting of NCME, San Francisco.

Hambleton, R.K., & Powell, S. (1983). A framework for viewing the process of standard-setting. Evaluation & the Health Professions, 6, 3-24.

Livingston, S.A., & Zieky, M. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, NJ: Educational Testing Service.

Popham, W.J. (1986, April). Preparing policymakers for standard-setting on high-stakes tests. Paper presented at the annual meeting of AERA, San Francisco.

van der Linden, W.J. (1982). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting. Journal of Educational Measurement, 19, 295-308.

Appendix A

Task Review Form

Task Review Form

Strategy: This form should be used with the group of judges to help the group reach a joint understanding of what minimum competency is for each task or objective. (In the form, the word "Task" is substituted for "Sub-Responsibility" for convenience.)

Each judge should determine the percent of times that a task or objective is to be accomplished with no or only a few minor errors. As a group, the judges should reach a compromise rating among their collective ratings.

Form:

Directions: Read each task in the role delineation statement (domain specification or objective) and determine the percent of times each task (objective) must be accomplished with no or only a few minor errors. For example, consider the following task:

Complete a standard order form for ordering office supplies

For this example, what percent of times that an order form is to be completed must the form be completed with no or only a few minor errors?

Task X. _____%

The response is __% of the times the order form must be completed with no or only a few minor errors.

Now, ask judges to look at the tasks in the role delineation profile.

What percent of times should each task be performed with no or only a few minor errors?

PES.7

Write a percent in the space provided.

1.____%	11.____%	21.____%	31.____%
2.____%	12.____%	22.____%	32.____%
3.____%	13.____%	23.____%	33.____%
4.____%	14.____%	24.____%	34.____%
5.____%	15.____%	25.____%	35.____%
6.____%	16.____%	26.____%	36.____%
7.____%	17.____%	27.____%	37.____%
8.____%	18.____%	28.____%	38.____%
9.____%	19.____%	29.____%	39.____%
10.____%	20.____%	30.____%	40.____%

Appendix B

Angoff Item Review Form
(Method A)

Angoff Item Review Form

Reviewer's Name:_____ Date:_____

Task (Objective) Statement: (insert the task objective number here)

This task objective must be performed _____% of the time with no or only a few errors.

I. Ask judges to think of a group of persons who are just able to meet this required level of performance for this task (objective).
The exam items below were prepared to measure this task (objective). What percent of the group of people that you are thinking about will be able to answer each exam item correctly? Write the Percent (between 0 and 100) for each exam item in the column labelled "Initial Percent."

Test Item	Initial Percent	Revised Percent
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%
___	_____%	_____%

II. When the judges in the work group have provided their initial ratings, ask them to compare their percents on an item by item basis. Also, review the scoring key. Identify the judges who have the highest and lowest percent for each exam item. If they are greatly different (about 20% points difference) then they should discuss why the percents were chosen. They do not have to reach a compromise. Only reconsider their own ratings when there are large differences. If they want to change their percents for any exam item, they should write a new percent in the Revised Percent column.

PES.7

Appendix C

Angoff Item Rating Form
(Method B)

Angoff Item Rating Form

Judge: _____ Exam: _____ Date: _____

Item	First Rating	Revised Rating	Item	First Rating	Revised Rating
1			21		
2			22		
3			23		
4			24		
5			25		
6			26		
7			27		
8			28		
9			29		
10			30		
11			31		
12			32		
13			33		
14			34		
15			35		
16			36		
17			37		
18			38		
19			39		
20			40		
Sub-total					

Continued -

Appendix D

Implementation Checklist for
the Angoff Standard-Setting Method

Implementation Checklist for
the Angoff Standard-Setting Method

Category/Activity	Date Task Completed
I. Initial Information	
1. Describe the purpose of the exam.	_____
2. Describe the importance of the exam.	_____
3. Consider if mastery/nonmastery decisions are required.	_____
4. Describe significant limitations of resources.	_____
5. Describe significant regulations and restrictions influencing the exam.	_____
6. Describe how the exam was developed.	_____
II. Standard-setting Meeting Preparation	
A. Background (Choose one of the options below.)	
Role Delineation Available	
1. (a) Prepare the role delineation statement.	_____
(b) Determine the appropriate level at which to obtain ratings.	_____
(c) Code exam items by the sub-responsibility (or level chosen) they measure.	_____
Domain Specifications Available	
1. Code exam items by the domain specs they measure.	_____
Objectives Available	
1. Code exam items by the objectives they measure.	_____
Exam Blueprint Available	
1. Code exam items to the cell of the exam blueprint they measure.	_____
2. Prepare a scoring key.	_____
3. Compile exam item statistics.	_____

PES.7

B. Meeting Site Selection

1. Choose a location for the meeting. _____
2. Insure sufficient workspace. _____
3. Arrange for coffee service and lunch. _____
4. Consider the incentive of a nice retreat site. _____

C. Judge Selection

1. Identify desired and required demographic and technical characteristics. _____
2. Determine the number of judges. _____
3. Determine if compensation is necessary. _____
4. Select judges. _____
5. Organize judges into work groups. _____

D. Preparation of Judges' Packages _____

III. Standard-Setting Meeting

A. Background

1. Order of work _____
2. History of the project _____
3. Rationale for setting the standard _____
4. Importance of the exam and the standard _____

B. Definition of a Minimally-Competent Professional

Method A

1. Review role delineation profile, domain specifications or list of objectives. _____
2. Instruct the judges in the use of the Task Review Form. _____
3. Have each judge complete the Task Review Form. _____
4. Reach compromise ratings. _____
5. Compile final listing of ratings. _____
6. Add ratings to the Angoff Item Review Forms. _____

Method B

1. Have judges review the exam blueprint. _____
2. Initiate a discussion about the minimally competent candidate; generate a list of skills. _____
3. Review the skills and revise. _____
4. Prepare a revised copy of the skills for judges. _____

C. Setting the Standard

1. Work with judges on a practice exercise. _____
2. Answer judges' questions about the review. _____
3. Distribute the Angoff Item Review Forms. _____
4. Separate the judges into work groups. _____
5. Have the judges complete the item reviews. _____
6. Aggregate the exam item ratings by work group. _____
7. Review ratings with judges. _____
8. Aggregate ratings for each exam item. _____
9. Determine the overall standard. _____
10. Review and discuss the standard with the judges. _____
11. Adjust the standard if necessary. _____

IV. Post Standard-Setting Meeting

A. Documentation of the Process

1. Outline the major steps taken. _____
2. Note any problems, inconsistencies or threats to the validity of the standard. _____

B. Distribute the final Documentation

1. Send copies of the documentation report to the judges. _____
2. Place a copy on file. _____